DOCUMENT RESUME

ED 034 965 AL 001 510

AUTHOR Gledhill, Donald F.; Wu, C. K.
TITLE Coding Chinese Characters on an Ordinary Typewriter.
PUB DATE 20 Aug 68
NOTE 4p.; Draft

EDRS PRICE EDRS Price MF-$0.25 HC-$0.30
DESCRIPTORS *Chinese, *Codification, *Computational Linguistics, Computer Storage Devices, Machine Translation, Romanization, *Typewriting, *Written Language
IDENTIFIERS *Lantran Chincode System

ABSTRACT

Presented in this paper is an illustration of the Lantran Chincode System for coding the Chinese language on ordinary English language typewriters or computer terminal keyboards. The key element of the Lantran coding is the use of the Pinyin phonetic romanization which has been officially standard in Communist China for the past 10 years. The end letters in the full coding for each character are the phonetic transcription of the pronunciation of that character in the Pinyin. Since many characters have the same pronunciation, the first part of the coding indicates as concisely as possible which particular character is intended. In general this means specifying the radical of the character and the number of additional strokes required, these two items together with the pronunciation being sufficient to correctly identify the written character in all but a few cases. The key board arrangement pictured is that of the Frieden Communications Terminal generating USASCII code for computer. (Authors/AMM)

ED034965

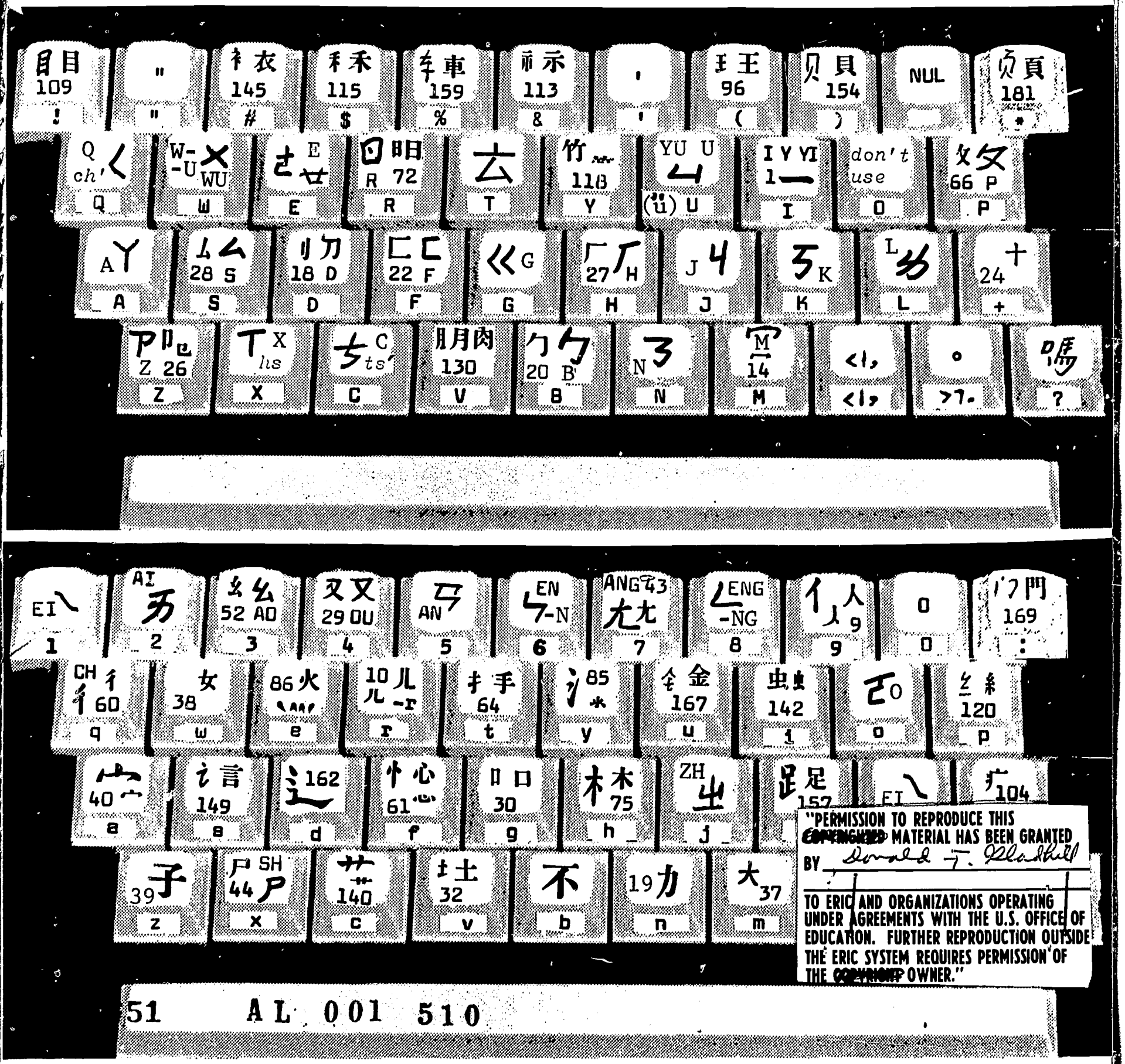
U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Coding Chinese Characters On an Ordinary Typewriter

Lantran Systems
Donald F. Gledhill
1617 Oramas Road
Santa Barbara, Calif. 93103

Chinese Language Consultant
C. K. Wu
Defense Language Institute
Monterey, Calif. 93940



"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Donald F. Gledhill
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

Keyboard arrangement is that of Friden Communications Terminal generating USASCII code for computer. Punctuation locations will vary on other typewriters.

COPYRIGHT RESERVED

Draft August 20, 1968

THE LANTRAN CHINCODE SYSTEM

For Ordinary English Language Typewriter or Computer Terminal Keyboard

Key element of the Lantran coding is the use of the Pinyin phonetic romanization which has been officially standard in Communist China for the past ten years. The end letters in the full coding for each character are the phonetic transcription of the pronunciation of that character in the Pinyin. Since many characters have the same pronunciation the first part of the coding indicates as concisely as possible which particular character is intended. In general this means specifying the radical of the character and the number of additional strokes required, these two items together with the pronunciation being sufficient to correctly identify the written character in all but a few cases.

To illustrate: The pronunciation LIANG applies to a dozen or so commonly used characters. By referring to the keyboard illustration it is apparent that LIANG can be coded by striking only three keys:

L = L
I = I
ANG = 7, so LI7 will be the code ending.

The first two elements will be the dictionary information of the radical, given in three digits, and the added strokes, given in one or two digits. Thus:

	Character	Radical	Strokes	Pinyin	Full Code
(measure)	量	166	5	LI7	1665LI7
(cold)	涼	085	8	LI7	0858LI7
(two)	兩	011	6	LI7	0116LI7
(bright)	亮	008	7	LI7	0087LI7
(vehicles)	輛	159	8	LI7	1598LI7
(grain)	糧	119	12	LI7	11912LI7
(good)	良	138	1	LI7	1381LI7

The end coding must be written in capital and lower case letters according to the location of the phonetic symbols on the keyboard illustration. Thus: DA for DA, Z for ZI and I8 for YING but Ko8 for KONG, x for SHI, and x6 for SHEN. (Capital N is used only for the initial sound.) The coding follows the Pinyin

particle 個 or 个 , with other characters pronounced GE requiring the longer code to distinguish them.

As the Lantran system comes into general use, particularly for computer-assisted translation from Chinese to English, short-coding possibilities will be developed until a substantial part of the language may be coded in Pinyin with conventionalized symbols identifying frequently used expressions and full coding required only for less common characters.

Editing and coding symbols: Punctuation marks written between characters have been reserved in the Lantran system for use in pre-editing. They indicate choices among translations. When typed into coding for the computer the marks result in a different output. They are also useful, however, when the coding dictionary is used ordinarily for study and translation.

- A dash between characters or between two or more units of coding indicates that they are to be joined for translation.

  Thus: 洋 - 火 The dash indicates that the two characters are not to be translated separately as foreign fire but in combination as matches.

' Single, double and triple quote marks (which may be called
" "as," "astu" and "astre") indicate additional translation
'" choices, chiefly for verbs and adverbs.

, Single, double and triple commas ("al, altu and altre")
,, indicate other translation choices, chiefly for nouns and
,,, adjectives, and for measures and other grammatical forms.

The above marks follow the character or coding they affect. The next mark goes ahead.

L This is the old Chinese sign 了 put to a new use. (It may be called "iya") It indicates that the character or more often the group of characters following has been short-coded. The effect is to greatly extend the range of phonetic coding. Thus 牠 0933TA, meaning it, is reduced to LTA (iya TA), distinguishing from the short-coded TA meaning he.

L,KEx is short-coding for 可-是 0302KE-0725x. The L, indicates that this KE-SHI is the syntactical one meaning but, distinguishing it from the characters also pronounced KE-SHI but meaning corrosion, guestroom, limestone, silicon, classroom, etc.

Radicals on Keyboard, coded by one stroke each													
1	I	22	F	32	v	52	3	85	y	118	Y	154	)
9	9	24	+	37	m	60	q	86	e	120	p	157	k
10	r	26	Z	38	w	61	f	96	(	130	V	159	%
14	M	27	H	39	z	64	t	104	;	140	c	162	d
18	D	28	S	40	a	66	P	109	!	142	i	167	u
19	n	29	4	43	7	72	R	113	&	145	#	169	:
20	B	30	g	44	x	75	h	115	$	149	s	181	*

in making distinctions among the CH sounds, thus q7 for CHANG 倘 QI4 for CHYOU 球, jW for ZHU 珠 and JIE for JIE 子 coded 039JIE, or zJIE, to distinguish from the radical main form 子, coded 39 or z. CW6 is CUN 寸 (ts'un in Wade) and Zo is ZUO 做

Radicals on the keyboard: In order to shorten the coding, English letters and symbols on the typewriter keyboard have been selected to represent the radicals for well over half of all the Chinese characters. This cuts the radical part of the coding from three strokes to one. So 凉 is coded y8LI7 instead of the full code 0858LI7, and 辆 is coded %8LI7 instead of the full code 1598LI7. A radical on the keyboard when used by itself for a word is coded by a single stroke: m for 037 大 as big, I for 001 一 as one and g for 口 030 as mouth. Radicals not on the keyboard may also be coded by just the radical number without the pronunciation--063, 110, 207, etc. 95W1 means radical 9 plus 5 strokes pronounced WEI. Radical 95, not on the keyboard is coded 095.

Short-coding: Chinese characters vary widely in frequency of use. Just as in English the expressions for is, does, has, I, she, with, go, today, this, that and so on make up much of the language. If a hundred or more such words are given the shortest possible coding, even with arbitrary combinations, frequent repetition will still make them easy to remember. The coding labor for any body of text will be reduced by hundreds of strokes with a gain in speed and efficiency. Unlike such rigid lists as the Chinese telegraph code, the Lantran system can profit by such short-coding, usually keeping all or part of the Pinyin pronunciation as a memory guide. For example 我们 is short-coded to WoM6, only four strokes. The short-coding GE is used for the